DOCUMENT RESUME

ED 449 177 TM 032 257

AUTHOR Brooks, Gordon P.; Barcikowski, Robert S.
TITLE The Precision Efficacy Analysis for Regression Sample Size Method.
PUB DATE 1999-04-00
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation; *Effect Size; Monte Carlo Methods; *Regression (Statistics); *Sample Size; *Selection
IDENTIFIERS *Precision (Mathematics)

ABSTRACT

The general purpose of this study was to examine the efficiency of the Precision Efficacy Analysis for Regression (PEAR) method for choosing appropriate sample sizes in regression studies used for precision. The PEAR method, which is based on the algebraic manipulation of an accepted cross-validity formula, essentially uses an effect size to determine the subject-to-variable ratio appropriate for the squared multiple correlation expected in a given study. An effort was made to determine how appropriate the sample sizes calculated by the PEAR method are for use with stepwise regression. A Monte Carlo analysis of precision efficacy rates was performed, manipulating effect sizes, predictors, and multicollinearity conditions, and using Turbo Pascal procedures to generate sample data. The PEAR method recommended sample sizes that provided reliable regression coefficients. Higher precision efficacy levels provided more stable coefficients. The use of the PEAR method in stepwise regression analyses proved less conclusive. For orthogonal predictors, the PEAR method did not fail, but as multicolinearity increased, the results were less impressive. Results suggest that for less multicolinear data, precision efficacy levels do not drop dramatically for stepwise analysis. Four appendixes contain figures illustrating the discussion. (Contains 11 tables and 56 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Running Head: PEAR Method

G. Brooks

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The Precision Efficacy Analysis for Regression Sample Size Method

ED 449 177

Gordon P. Brooks

Robert S. Barcikowski

Ohio University

TM032257

Paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada (1999, April)

BEST COPY AVAILABLE

THE PRECISION EFFICACY ANALYSIS FOR REGRESSION SAMPLE SIZE METHOD

"I have so heavily emphasized the desirability of working with few variables and large sample sizes that some of my students have spread the rumor that my idea of the perfect study is one with 10,000 cases and no variables. They go too far." (Cohen, 1990, p. 1305). Although Darlington (1990), among others, has noted that the best rule for choosing sample sizes is simply that more is better, 10,000 may be just a couple more than typically are necessary. Indeed, for both statistical and practical reasons, researchers should choose for their sample size "the smallest number of cases that has a decent chance of revealing a significant relationship if, indeed, one is there" (Tabachnick & Fidell, 1989, p. 129).

When generalizability is the primary concern, as it is when regression is used to develop prediction models, this concept translates as the smallest sample that will provide the required reliability of results across multiple samples. Especially in multiple linear regression, which is used for many purposes, necessary sample size depends heavily on the goals and design of the analysis. Consequently, the selection of adequate and appropriate sample sizes is not always an easy matter in regression.

Several methods currently exist to help researchers choose sample size, including conventional rules, statistical power methods, and cross-validation methods. Unfortunately, because of difficulties and contradictions among these various methods, sample size selection in multiple regression has been problematic. For example, how does one reconcile the difference among Cohen's (1988) statistical power method that recommends 48 subjects, Park and Dudycha's (1974) method that advises 93 subjects, and Stevens' (1996) 15:1 subject-to-predictor ratio that suggests 60? See Table 1 for several such discrepancies.

The general purpose of this study was to examine the efficiency of the Precision Efficacy Analysis for Regression (PEAR) method for choosing appropriate sample sizes in regression studies used for prediction. The PEAR method, which is based on the algebraic manipulation of an accepted cross-validity formula, essentially uses an effect size to determine the subject-to-variable ratio appropriate for the squared multiple correlation expected in a given study. For example, using one set of criteria at an effect size of expected $\rho^2 = .40$, the PEAR method suggests a subject-to-variable ratio of approximately 15:1; but with the same criteria at an

expected ρ^2 of .20, the number of subjects required per variable increases to nearly 38:1. See Table 2 for sample sizes recommended by the PEAR method for several other criteria.

Theoretical Perspectives

When researchers are most interested in testing the statistical significance of either a sample multiple correlation or particular independent variables, several statistical power sample size methods exist for those purposes (e.g., Cohen, 1988; Cohen & Cohen, 1983; Gatsonis & Sampson, 1989; Kraemer & Thiemann, 1987; Milton, 1986). Unfortunately, statistical power to reject a regression null hypothesis does not provide information about the number of subjects needed to obtain the stable, meaningful regression coefficients required for prediction. Therefore, choosing a sample size based on statistical power may not ensure that a regression function will generalize to other samples from the target population, which is the crucial factor in determining the validity of regression models to be used for prediction.

Alternatively, conventional rules have evolved that are based on the premise that with a large enough ratio of subjects to predictors (e.g., 10 or 15 subjects for each predictor), the sample regression coefficients will be reliable and will closely estimate the true population values. Unfortunately, because most of these rules lack any measure of effect size, they can only be effective at specific effect sizes—which may not be appropriate for any given study. For example, a 15:1 subject-to-variable ratio is acceptable only if the population squared multiple correlation is over .40; otherwise, as the true squared multiple correlation decreases, expected cross-validity shrinks so much as to make the prediction model worthless (Brooks & Barcikowski, 1995).

Park and Dudycha (1974) were among the first to define mathematically a sample size method using a random model, cross-validation approach. Unfortunately, they published tables that were limited to only a few possible combinations of squared correlation and number of predictors; also, their math is too complex for many researchers to derive the information needed for the cases not tabulated. Darlington (1990) has provided two precision methods, but one provides recommended sample sizes for only the validation sample (i.e., not the original derivation sample) and the other provides sample sizes for better estimation of the true population correlation rather than the cross-validity coefficient.

Due to the lack of an adequate method to determine sample sizes that ensures some

measure of cross-validity, the PEAR method was developed. The primary goal of the PEAR method is to reduce the upward bias of R^2, thereby enhancing the cross-validity potential of the model so that results are less likely to be sample specific. In a sense, the PEAR method can be viewed as cross-validation in reverse. That is, instead of determining by how much the sample R^2 will shrink due to the sample size, the PEAR method determines how large a sample is required to keep R^2 from shrinking too much. The theory underlying the PEAR method for sample size selection is that the researcher, knowing that cross-validation is likely to cause shrinkage in R^2, can set a limit as to the amount of shrinkage expected to occur. The concepts of cross-validity shrinkage, precision efficacy, proportional shrinkage, effect size, and shrinkage tolerance serve as the foundation for using the PEAR method of sample size selection to, in Stevens' terms, "keep the shrinkage fairly small."

Cross-Validity Shrinkage

"Although we may determine from a sample R^2 that the population R^2 is not likely to be zero, it is nevertheless not true that the sample R^2 is a good estimate of the population R^2" (Cohen & Cohen, 1983, p. 105). While most questions concerning explanation, description, and causal analysis require an adjusted R^2 estimate of ρ^2 (such as the common R_A^2 formula most often attributed to Wherry), most problems of prediction are concerned primarily with cross-validity. From a generalizability viewpoint, an insufficient sample leads to results that, even though maybe statistically significant, may apply only to the current sample and will not be useful or practical for application to other samples. As Herzberg (1969) noted, "in applications, the population regression function can never be known and one is more interested in how effective the *sample* regression function is in *other* samples" (p. 4). Therefore, researchers must use and report strategies that evaluate the replicability of their results; the best way to gauge this generalizability is through an estimate of cross-validity. The squared cross-validity coefficient, ρ_C^2, is considered to be the squared multiple correlation between the actual population criterion values and the scores predicted by the sample regression equation when applied either to the population or to another sample (Cattin, 1980b; Huberty & Mourad, 1980; Kennedy, 1988; Schmitt, Coyle, & Rauschenberger, 1977).

Cross-validity correction formulas, symbolized by R_C^2, which are based on estimates of the mean squared error of prediction (Darlington, 1968; Herzberg, 1969), provide more accurate

estimates of ρ_C^2 than does R^2. Formula methods of cross-validity are often preferred to empirical cross-validation (e.g., data-splitting) so that the entire sample may be used for model-building. Indeed, several formula estimates have been shown superior, or at least equivalent, to empirical cross-validation techniques (Cattin, 1980a, 1980b; Drasgow, Dorans, & Tucker, 1978; Kennedy, 1988; Morris, 1981; Rozeboom, 1978; Schmitt, Coyle, & Rauschenberger, 1977). Many such cross-validity formulas have been proposed (e.g., Browne, 1975; Darlington, 1968; Herzberg, 1969; Lord, 1950; Nicholson, 1960; Rozeboom, 1978; Stein, 1960).

When shrinkage is calculated through the use of a cross-validity formula, any finite sample size will result in a cross-validity estimate that is smaller than the sample R^2. Similar conceptually to Cronbach's reliability coefficient alpha, cross-validity formulas attempt to estimate the average of all possible empirical cross-validations (Wherry, 1975). For example, using the random model cross-validity estimate developed independently by Stein (1960) and Darlington (1968), $R_C^2 = 1 - [(1 - R^2)(N - 1)(N - 2)(N + 1)]/[(N - p - 1)(N - p - 2)(N)]$, a researcher who calculates a sample $R^2 = .400$ with 60 subjects and 4 predictors might calculate the sample squared cross-validity as $R_C^2 = .297$ (note that the Wherry R_A^2 is .356 for these conditions). This cross-validity estimate implies that the researcher might be more likely to explain 30%, not 40%, of the variance of the criterion when applying the sample regression function to future samples.

Precision Efficacy

Precision efficacy (PE) describes how well a regression model is expected to perform when applied to future subjects relative to its effectiveness in the derivation sample. The formal definition of precision efficacy is $PE = R_C^2 / R^2$, where R^2 is the sample coefficient of determination and R_C^2 is the sample cross-validity estimate. Because they desire regression models that generalize well to other samples, researchers who develop prediction models hope to limit shrinkage as much as possible relative to the sample R^2 value they attained.

Using an example from Stevens (1996, p. 100), 62% shrinkage from a sample $R^2 = .50$ to $R_C^2 = .191$ occurs with a sample size of 50; but if the sample size had been 150, there would have been only 16% shrinkage from the same $R^2 = .50$ to $R_C^2 = .421$. The precision efficacy in the first case would be $.191/.50 = .382$ and in the second case $PE = .842$. Consequently, even if the R^2 value was significant in the first case, the results may not be expected to perform

well enough for the model to be useful with future samples. Larger precision efficacy values imply that a regression model is expected to generalize better for future samples.

Proportional Shrinkage. Proportional shrinkage (*PS*) is the amount of shrinkage relative to R^2 that occurs after a cross-validity estimate, R_C^2, is calculated from the data. Proportional shrinkage is calculated by $PS = (R^2 - R_C^2) / R^2$. The precision efficacy of the regression equation, and therefore an estimate of the model's generalizability, also can be computed as $PE = 1 - PS$. For example, if sample $R^2 = .50$ and $R_C^2 = .26$, the proportional shrinkage for that regression model can also be calculated as $PS = (.50 - .26)/.50 = .48$. Proportional shrinkage of .48, and therefore $PE = .52$, suggests limited generalizability for the regression model because the R^2 value shrank by almost half.

Effect Size

In multiple regression research, perhaps the most common effect size is the squared multiple correlation, R^2. Effect size enables a researcher to decide a priori not only what size relationship will be necessary for statistical significance, but also what relationship should be considered for practical significance (Hinkle & Oliver, 1983). Light, Singer, and Willett (1990) offered as a starting point that this effect size should be "the minimum effect size you consider worthy of your time" (p. 194). For example, because under 10% explained variance may not provide any new knowledge in the field, a researcher may choose a minimum practical effect size of 20%. In multiple regression, however, the researcher must remember the effects of shrinkage—if a researcher chooses 20% explained variance (i.e., $R^2 = .20$) as a minimum practical effect worthy of study, that researcher does not want a corrected sample estimate (e.g., R_A^2 or R_C^2) to be .05.

There are three basic strategies for choosing an appropriate effect size: (a) use effect sizes found in previous studies or meta-analysis, (b) decide on some minimum effect that will be practically significant, or (c) use conventional small, medium, and large effects such as those defined by Cohen (1988). No matter how it is chosen, effect size must be chosen a priori. In many cases, the researcher may have some basis for deciding the smallest correlation that would be interesting to find, based perhaps on experience or prior research. In other cases, however, researchers may need to rely on intuition or other means by which to choose an effect size. For example, Stevens (1986) has suggested that $\rho^2 = .50$ is a reasonable guess for social science

research; Rozeboom (1981), however, believed $\rho^2 = .50$ to be an upper limit. Indeed, because an effect of $\rho^2 = .25$ seems unreasonably large to Schafer (1993), he recommended that it serve as an upper limit only as a last resort, when no other rationale is available. Light, Singer, and Willett (1990) echoed Schafer: "meta-analyses often reveal a sobering fact: effect sizes are not nearly as large as we all might hope" (p. 195).

The relationship between effect size and sample size. Stevens (1996) has emphasized that the magnitude of the population squared multiple correlation, ρ^2, "strongly affects how many subjects will be needed for a reliable regression equation" (p. 125). For example, Stevens (1996, p. 125) demonstrated that "more than 15 subjects per predictor will be needed to keep the shrinkage fairly small" if .40 is used as R^2 in the Stein cross-validity formula, but that fewer will be needed if $R^2 = .70$. Similarly, Huberty (1994) noted that based on analysis of shrinkage results that "it is perhaps clear that the magnitude of R^2 should be considered in addition to N/p ratios when assessing the percent of shrinkage of R^2 that would result in the estimation process. That is, a general rule of thumb for a desirable N/p ratio (say, 10/1) may not be applicable across many areas of study" (p. 356). Indeed, all methods that account for effect size agree: as effect size decreases, sample size must increase proportionately (e.g., Cohen, 1988; Darlington, 1990; Milton, 1986; Park & Dudycha, 1974; Gatsonis & Sampson, 1989). Therefore, the first task in any sample size analysis generally is regarded to be the identification of the expected magnitude of the multiple correlation in the population.

Shrinkage Tolerance

Simply put, shrinkage is the size of the decrease in the sample R^2 when an appropriate cross-validity formula is applied. Shrinkage tolerance, an a priori definition of acceptable shrinkage, can be defined mathematically as $\epsilon = R^2 - R_C^2$. Shrinkage tolerance can be considered either absolute or relative. In an absolute sense, ϵ can be set to a specific value regardless of the effect size expected in a given study. That is, no matter what R^2 is to be used, the researcher may wish that the expected shrinkage be within .10 of the sample R^2 value. For example, if R^2 is expected to be near .50 and the researcher has chosen $\epsilon = .10$, R_C^2 will be expected to be near .40; but if R^2 is expected to be near .35, the researcher is willing to accept .25 for the expected shrunken R_C^2 value when ϵ is set to .10.

In a relative sense, the formula for calculating precision efficacy can also be written as

$PE = 1 - \epsilon/R^2$. For example, setting the predetermined acceptable shrinkage level at $\epsilon = .2R^2$ provides precision efficacy of .80. To provide a numerical example, if the population ρ^2 is thought to be .50 and ϵ is set at $.2R^2$, the sample R^2 is expected to shrink only by 20% to $R_C^2 = .40$ and hence precision efficacy of .80; whereas, if expected R^2 is near .35, R_C^2 would be expected near .28—again $PE = .80$. Or if ϵ is set at $.3R^2$, a sample R^2 of .50 will be expected to shrink by 30% to $R_C^2 = .35$, a PE of .70.

Solving $PE = 1 - \epsilon/R^2$ for ϵ and replacing R^2 with an a priori R_E^2 results in the formula $\epsilon = R_E^2 - (PE \times R_E^2)$, where R_E^2 is the expected sample R^2 effect size value chosen by the researcher. Using this formula, a specific level of precision efficacy can be set a priori to determine the acceptable shrinkage tolerance to use in selecting an adequate sample size. For example, if the researcher wishes to obtain a cross-validity estimate expected to be not less than 80% of the sample R^2, a priori precision efficacy would be .80. If the expected sample R^2 is though to be $R_E^2 = .50$, then the shrinkage tolerance can be found by substituting the appropriate values the equation for ϵ. That is, shrinkage tolerance would be found a priori for this example by $\epsilon = .50 - (.80 \times .50) = .10$.

It should be noted that in the course of the development of the PEAR method, because R^2 is a positively biased estimator of both ρ^2 and ρ_C^2 such that $E(R^2) > \rho^2 > \rho_C^2$, it was determined that a slight modification to the shrinkage tolerance formula performs better when an estimate of ρ^2 is more readily available than an estimate of R^2 (Brooks, 1998b). This modified ϵ is calculated by $\epsilon = \rho_E^2 - (PE - .1PS)\rho_E^2$, where $PS = 1 - PE$ and ρ_E^2 is the estimated population ρ^2 value (e.g., R_A^2 found in previous research or through meta-analysis). Using the same example from above results in the following: $\epsilon = .50 - ([.80 - .1(.20)] \times .50) = .11$.

The PEAR Method

The PEAR method sample size formula was developed based on a cross-validity formula by Lord (as cited in Uhl & Eisenberg, 1970): $R_C^2 = 1 - (N + p + 1)(1 - R^2)/(N - p - 1)$, where N is sample size, p is the number of predictors, and R^2 is the actual sample value. Uhl and Eisenberg (1970, p. 489) found this "relatively unknown formula" (their interpretation of Lord, 1950, differs from others) to give accurate estimates of "cross-sample" shrinkage, regardless of sample size and number of predictors. Algebraic manipulation of the Lord formula to solve for sample size yields the Precision Efficacy Analysis for Regression method sample

size formula for multiple linear regression:

$$N = (p + 1) \times \frac{(2 - 2R_E^2 + \epsilon)}{\epsilon}$$

where p is the number of predictors. R_E^2 is the a priori expected sample R^2, and ϵ is an acceptable a priori amount of expected shrinkage. The R_E^2 serves as an effect size (note that when using an estimated ρ^2, the appropriate ϵ formula should be used and that ρ_E^2 should be used in place of R_E^2). Shrinkage tolerance allows researchers to decide how closely to estimate ρ_C^2, either as an absolute amount of acceptable shrinkage (e.g., $\epsilon = .05$), a proportional decrease (e.g., $\epsilon = .2R_E^2$, which represents shrinkage of 20% from R_E^2 to $R_C^2 = .8R_E^2$), or using the ϵ formula described above. It is also worth nothing that Brooks and Barcikowski (1995) determined that the total number of variables, $(p + 1)$, performs better in the PEAR method than does the number of predictors. A derivation of the PEAR method formula has been included in Appendix A.

If a researcher wanted an R_C^2 estimate to be at least 87% of the expected sample R_E^2 of .53 with four predictors, the researcher would set PE to .87 and calculate $\epsilon = .53 - (.87 \times .53) = .069$. These values would then be substituted into the PEAR method formula to calculate the necessary sample size as 73.12. Therefore, at least 74 subjects should provide a large enough sample so that R_C^2 is expected to be greater than .46, which is 87% of the assumed ρ^2 of .53. Another example illustrates that if PE is desired to be .80 when using an estimated ρ^2, shrinkage tolerance is calculated as $\epsilon = .22\rho_E^2$ and the PEAR method formula simplifies slightly to $N \geq (p + 1)(2 - 1.78\rho_E^2)/(.22\rho_E^2)$.

Development of the PEAR Method

Early research in its development found the PEAR method to be superior to statistical power methods (Cohen, 1988; Gatsonis & Sampson, 1989), conventional rules (Green, 1991; Pedhazur & Schmelkin, 1991; Stevens, 1996), and cross-validity methods (Park & Dudycha, 1974; Sawyer, 1982) in reliably and accurately limiting cross-validity shrinkage to given acceptable a priori levels (Brooks & Barcikowski, 1995). Specifically, using an accuracy interval of $.75 \leq PE \leq .85$, the PEAR method provided accurate precision efficacy rates (i.e., actual PE within .05 of nominal $PE = .80$) in all 20 conditions where expected R^2 approximated true ρ^2 (see Appendix B). The accuracy of the other regression sample size methods was low relative to

the PEAR method, with none of these methods accurate at $PE = .80$ for more than five of the 20 conditions. Furthermore, whereas the PEAR method provided consistent results across all conditions, the other methods varied considerably in actual PE rates across both the number of predictors and expected R^2 values.

Brooks (1998a) reported that, using a bias accuracy criterion of $|E(PE) - PE| \leq .1PS$ where $PS = 1 - PE$, the PEAR method maintained the perfect accuracy it had shown for $PE = .80$ when the precision efficacy rate was lowered to $PE = .70$ and also 87.5% accuracy for $PE = .60$. The $PE = .80$ level of precision efficacy was determined to be about 20% more efficient (i.e., standard errors which on average were 20% smaller) than the $PE = .70$ level, which in turn was about 14% more efficient than the $PE = .60$. Additionally, the results showed this pattern of Relative Efficiency to hold true no matter what level of multicollinearity was present in the predictor sets.

Because previous work has focused on the effects of sample size on the correlation statistics for the full regression model, the current report examines impact of the PEAR method sample sizes on the variance of the regression coefficients. First, does the PEAR method recommend sample sizes that enable the derivation of reliable regression coefficients (that is, coefficients with small standard errors)? In order to examine the stability of the coefficients, the standard errors of the coefficients (SE_{b_j}) are of primary interest. One would expect that a model based on a proper sample size will provide more reliable regression weights and therefore predict better for future subjects. Second, despite the well-known disadvantages of stepwise regression, it is a common method used by researchers, particularly as a means by which to handle multicollinearity (Breiman, 1995; Huberty, 1989). Therefore, an effort was made to determine how appropriate the sample sizes calculated by the PEAR method are for use with stepwise regression.

Method

The cross-validational efficiency of sample size methods can be assessed analytically to some extent. Once a sample size has been chosen via any sample size method for a given number of predictors and a given expected ρ^2, cross-validity can be estimated. For example, once the number of predictors is set at four and ρ^2 is assumed to be .25, the sample required by the PEAR method at $PE = .80$ is 142. Using these values in the Stein-Darlington R_C^2 formula

gives an R_C^2 of .199, or 80% of the original ρ^2 value. Comparisons have been made in this way for several sample size methods in Table 3.

However, several elements of the current study did not lend themselves to such analysis. Therefore, a Monte Carlo analysis of precision efficacy rates was performed. The three PEAR method a priori precision efficacy levels of .60, .70, and .80 (which correspond to squared cross-validity estimates expected to be at least 60%, 70%, and 80% of the sample R^2 values, respectively) were considered to be individual methods for the analysis. That is, sample sizes were calculated using these *PE* levels with the PEAR method. Because the PEAR method has been shown previously to be superior to other regression sample size methods, only the *15:1* subject-to-predictor ratio was included for the sake of comparison. Comparisons of the varying precision efficacy levels of the PEAR method helped to determine the effects of larger and smaller sample sizes on the regression coefficients.

Because a variety of factors may influence precision efficacy, three factors were manipulated to comprise the testing situations for the study. First, three effect sizes that represent simultaneously the estimated population squared multiple correlation ρ_E^2 and the true population ρ^2 were set at: .10, .25, and .40. The numbers of predictors used to define the models in this study were 3 predictors (i.e., 4 variables including the criterion), 7, 11, and 15 predictors. Finally, four multicollinearity conditions were explored in the study: (1) *extensive* multicollinearity was defined as over one-half of the predictors with $VIF_j > 5.0$, (2) *moderate* multicollinearity was defined as one-quarter of the predictors involved in such a multicollinear relationship, (3) for all predictors in the *non-multicollinear* condition, $VIF_j < 3.0$, and (4) the correlation matrix for the *orthogonal* condition contained zero correlations among all predictors.

A Turbo Pascal program was created for an original algorithm used to create 48 population correlation matrices to meet the above criteria required by this study (explained in Brooks, 1998b). These correlation matrices, some of which can be seen in Appendix C, were treated as population correlation matrices from which multivariate normal data were generated for each sample in the study. Turbo Pascal procedures were developed to generate sample data through a process that converted uniformly distributed pseudorandom numbers created by the L'Ecuyer (1988) combined multiplicative congruential generator (translated from Press, Teukolsky, Vetterling, & Flannery, 1992) into multivariate-normally distributed data using the

Box-Muller transformation (adapted from Press, Flannery, Teukolsky, and Vetterling, 1989) and the Cholesky decomposition (adapted from Nash, 1990). Finally, these procedures were incorporated into a Turbo Pascal program that performed the Monte Carlo simulation with 10,000 iterations. The program was run as a DOS application under Windows 95 on a computer equipped with an Intel Pentium-MMX 133MHz processor. Double precision floating point variables were used, providing a maximum possible range of values between 5.0×10^{-324} to 1.7×10^{308}, stored with 15 to 16 significant digits.

Data Analysis Procedures

During program execution, several statistics were computed and recorded. For each sample, the program performed a standard multiple linear regression analysis based on algorithms provided in Barcikowski (1980) and a stepwise analysis based on Jennrich (1977). The program first calculated the necessary information from the full-model regression for each sample (e.g., *PE*, R^2, Wherry R_A^2, Stein R_C^2, β_j, SE_{b_j}). Both R_A^2 and R_C^2 were set equal to zero when they were negative, as recommended by Cohen and Cohen (1983) and Darlington (1990). These data were averaged over the number of iterations for each condition. Finally, counts were made for several statistics regarding their significance or accuracy. For example, statistical significance at $\alpha = .05$ was tested for both the full regression model and the regression coefficients, as was the accuracy of *PE* and R_C^2. Similar statistics were collected for the stepwise analyses, with appropriate adaptations such as $R_{C(p)}^2$ and $R_{C(k)}^2$, which estimate cross-validity for the total number of predictors and only the number of predictors in the final model, respectively.

In addition to these raw statistics, the appropriate calculations were made and data were collected as required for bias, root mean squared error (*RMSE*), Relative Efficiency, statistical power, and the standard deviations of several key estimates. Statistical bias is defined as the difference between the population value ρ^2 and the expected value of its estimate: $Bias = E(\hat{\theta}) - \theta$, where θ is the population parameter and $E(\hat{\theta})$ is the expected value of the sample statistic or an average of the statistic over infinite samples (Drasgow, Dorans, & Tucker, 1979; Kromrey & Hines, 1995; Mooney, 1997).

The root mean squared error (*RMSE*) provides an indication of the statistic's variability. Mean squared error is the average of the squared differences between the population parameter and its estimate for each sample. *RMSE*, then, is the square root of the mean squared error for

the given statistic: $RMSE(\hat{\theta}) = \sqrt{\Sigma(\theta - \hat{\theta}_i)^2/n}$, where θ is the known population parameter (as set in the computer algorithm), $\hat{\theta}_i$ is the estimate of that parameter obtained in sample *i* of the Monte Carlo simulation, and *n* is the total number of samples taken in the Monte Carlo study (Darlington, 1996; Drasgow, Dorans, & Tucker, 1979; Kennedy, 1988; Mooney, 1997). Mooney (1997) defined Relative Efficiency as the ratio of two *RMSE* values, multiplied by 100 to convert it to a percentage: $Relative\ Efficiency = 100 \times RMSE(\hat{\theta}_A)/RMSE(\hat{\theta}_B)$, where $\hat{\theta}_A$ and $\hat{\theta}_B$ are two different estimates the same parameter (Mooney, 1997). Values under 100 would indicate the superiority of estimator $\hat{\theta}_A$ (i.e., $\hat{\theta}_A$ with smaller *RMSE*).

In order to examine the stability of the coefficients, the standard errors were examined in order to determine how reliable the estimates were for each method. For the purpose of comparing sample size methods, the Relative Efficiency of the coefficients was examined. Several of the analyses could not be performed for the *15:1* ratio because no a priori precision efficacy rate could be fixed for this method.

Finally, it should be noted that the study was carried out from certain perspectives, which implied specific delimitations. That is, this study applied to standard ordinary least squares regression analysis with all predictors entered simultaneously in the full-model case. Also a random model perspective was assumed, where both the predictors and the criterion were sampled together from a joint multivariate normal distribution. The random model is often more appropriate for educational researchers and social scientists because they frequently measure random subjects on predictors and criterion simultaneously and therefore are not able to fix the values for the independent variables. Also, the current study considered only sample sizes required for multiple linear regression used to develop prediction models, one of the most common and most important uses of regression equations in the social sciences (Huberty, 1989; Weisberg, 1985). Consequently, the focus of this study is on the determination of sample sizes for the generalizability of prediction equations, not the power of statistical tests for null hypotheses concerning multiple correlation or the regression coefficients.

Results and Discussion

The PEAR method recommended sample sizes that provided reliable regression coefficients. More specifically, higher *PE* levels provided more stable coefficients. For the conditions with three predictors, Table 4 provides the standard errors of the coefficients for the

four sample size methods; similarly, Table 5 provides this information for seven predictor models. These tables show that the precision efficacy levels that recommended larger samples consistently resulted in smaller standard errors of the coefficients, regardless of the number of predictors or effect size. Although the problem of multicollinearity was not cured by the PEAR method, higher levels of precision efficacy do indeed help alleviate the effects. The results showed similar patterns for the 11 and 15 predictor cases as well.

Table 6 provides the relative efficiency of the methods compared for all numbers of predictors, all multicollinearity levels, and all effect sizes. For this table, the standard errors for the individual predictors were used for comparison because, for unbiased estimates such as the regression coefficients, *RMSE* approximates the standard error. To create Table 6, the relative efficiency of each predictor was calculated and then those values were averaged for the predictor set. It would not have been appropriate to average the results for Table 6 across predictors if the results had not been so consistent. For example, in Table 6 for $p = 3$ at $\rho^2 = .40$ in the orthogonal condition, the relative efficiency of the $PE = .80$ level as compared to $PE = .70$, represented as $RMSE(.80)/RMSE(.70)$, is shown to be 80.8%. Using the values from Table 4, it can be determined that for $p = 3$ at $\rho^2 = .40$ in the orthogonal condition, the relative efficiency for coefficients 1 was 80.9% (.102/.126); similarly, relative efficiency for coefficient 2 can be calculated to be 81.7% and for coefficient 3 at 79.6%.

There is a striking similarity between the relative efficiency statistics in Table 6 and those found by Brooks (1998a) for the correlation statistics. Specifically, the relative efficiency statistics show that, regardless of multicollinearity level, the magnitude of the standard errors of the coefficients from the $PE = .80$ level were, on average, about 19% or 20% smaller than those from the $PE = .70$ level. Similarly, Relative Efficiency comparisons of the $PE = .70$ and $PE = .60$ levels showed $PE = .70$ to be approximately 13% or 14% more efficient in terms of standard errors. Graphically, the distribution of one coefficient, which was involved in extensive multicollinearity, has been provided as Appendix D.

Stepwise Regression

The use of the PEAR method in stepwise regression analyses provided less conclusive results. For orthogonal predictors, the PEAR method did not fail; unfortunately, as multicollinearity increased, the results were less impressive. However, stepwise regression did

seem to help manage multicollinearity better than standard, full model regression. The average standard errors for the coefficients from the stepwise solutions were smaller than their full model counterparts. That is, when a multicollinear coefficient was kept in the final model, but others with which it correlated were removed, it usually was more precise due to smaller standard error.

Table 7 provides average precision efficacy rates for the stepwise analyses performed in the study. Table 7 shows that the method for calculating stepwise R^2_C using the total number of predictors in the full model, $R^2_{C(p)}$, tended to result in precision efficacy below that of the full model; however, the method by which the stepwise R^2_C was calculated using the number of predictors in the final model, $R^2_{C(k)}$, usually resulted in estimates above the full model R^2_C. That the $PE_{R^2_{C(k)}}$ values are larger than $PE_{R^2_{C(p)}}$ indicate that the two R^2_C estimates differ. For example, the average $R^2_{C(p)}$ value for the orthogonal $\rho^2 = .40$ at $PE = .80$ with three predictors was .337, but $R^2_{C(k)} = .350$. Examination of Table 8, which presents that precision efficacy based on $R^2_{C(p)}$, illustrates that stepwise precision efficacy was impacted by multicollinearity in stepwise analyses, even though neither PE nor R^2_C were affected by multicollinearity in the full model. Interestingly, Table 7 and Table 8 together indicate that the orthogonal stepwise PE rates based on $R^2_{C(p)}$ did not differ drastically from the full model counterparts; that is, these stepwise PE rates for the orthogonal condition often fell within the accuracy interval defined for the full model.

Table 9 provides the cross-validity estimates for $R^2_{C(p)}$ across the four multicollinearity conditions. In most cases, the estimate decreased further as multicollinearity became more extensive. In all cases, both the moderate and extensive multicollinearity conditions resulted in lower $R^2_{C(p)}$ estimates than were obtained in the orthogonal condition. The relative efficiency for $R^2_{C(p)}$ did not show a pattern similar to that shown by R^2_C in the full model situation; that is, the results varied considerably depending on the level of multicollinearity in the predictor set.

The results suggest that for less multicollinear data, precision efficacy levels do not drop dramatically for stepwise analyses. Table 8 indicates that more extensive multicollinearity requires larger samples to attain higher precision efficacy rates. That is, although the orthogonal and non-multicollinear conditions provide PE rates nearly as large as their a priori full model values, more extensive multicollinearity lowers the actual PE rates—sometimes substantially. However, for the $\rho^2 = .10$ effect size, which normally requires larger samples, the reduction in

PE rates in not nearly so severe as for the ρ^2 = .40 effect size (see Table 8).

The standard errors of the coefficients calculated and displayed in Table 10 were averages only for the samples in which the predictor was included in the final model. Table 10 reveals that in the stepwise situation, multicollinearity also caused the coefficients to become less stable. Table 11 shows that no convenient patterns of relative efficiency were present for the standard errors of the coefficients from the stepwise solutions beyond the orthogonal condition. However, at higher sample sizes, the standard errors from the orthogonal stepwise models were nearly equivalent to the full model (e.g., Table 5 versus Table 10).

Additionally, the larger relative decrease of the *PE* = .70 and the *PE* = .60 levels of precision efficacy from the orthogonal to the extensive multicollinearity conditions may indicate that higher *PE* levels are more appropriate for stepwise analyses. For example, from the orthogonal to extensive multicollinearity conditions, the actual *PE* rates for *PE* = .80 at ρ^2 = .40 with seven predictors decrease from .783 to .627 (or by 0.156); there was a decrease of 0.231 for *PE* = .70 and 0.254 for *PE* = .60 (see Table 8). Perhaps an a priori *PE* value of .85 or .90 would abate this decrease even more than the .80 level does. However, it certainly does not appear that arbitrary doubling of sample size is the proper solution to stepwise regression sample sizes (as is recommended by Tabachnick and Fidell, 1989, for example).

Because stepwise models usually result in slightly lower sample R^2 values, perhaps a reduction in the expected R_E^2 value, which in turn would result in a larger sample size, would be more appropriate. For example, fewer predictors in the final models as multicollinearity increased resulted in smaller average sample R^2 values: at *PE* = .80, ρ^2 = .40, and seven predictors, the orthogonal R^2 was .412, the non-multicollinear R^2 was .407, the moderately multicollinear R^2 was .360, and the extensively multicollinear R^2 was .307 (in contrast, the full model R^2 was very nearly .434 for each of the four multicollinearity levels).

Indeed, that there were fewer predictors in the final model also often reduced the standard error of prediction for the final model, as compared to the full model (Brooks, 1998b). Therefore, because the standard errors of the coefficients were usually smaller than for the full model, stability (relative to the full model) of the stepwise solution appears to be less a problem than whether the best theoretical model was chosen. But when multicollinearity is a population condition, "it matters little as far as prediction is concerned which of the variables involved in the

multicollinearity is removed from the estimated model" because the multicollinearity is always expected to be there (Mason, Gunst, & Webster, 1975, p. 289).

Finally, it must be recognized that stepwise analyses are complicated by the choice of $R^2_{C(p)}$ versus $R^2_{C(k)}$ for cross-validity. Because $R^2_{C(k)}$ was found to be liberal for many of the situations (cf. Cohen & Cohen, 1983; Derksen & Keselman, 1992), $R^2_{C(p)}$ was chosen for most of the analyses in the present study. Derksen and Keselman (1992) and Cohen and Cohen (1983) have recommended that an adjusted R^2 value calculated based on the full p predictors, $R^2_{A(p)}$, is better than $R^2_{A(k)}$ calculated by using only the number of predictors (k) in the final model. Derksen and Keselman found that although $R^2_{A(k)}$ was certainly a better estimate of the stepwise ρ^2 than R^2, it overestimated the population value for many subject-to-predictor combinations. It is interesting to note that as of version 7.0, SPSS uses stepwise regression summary statistics, including Adjusted R Square, that are based on the number of predictors that are "currently entered in the equation" (SPSS Inc., 1996, p. 434). However, current results do not support the notion that this issue has been decided.

Scientific and Educational Importance

The primary goal of Precision Efficacy Analysis for Regression is to provide a means by which the researcher can assess the prediction potential (i.e., generalizability) of a regression model relative to its performance in the derivation sample. As Cohen (1990) stated, "the investigator is not interested in making predictions for that sample—he or she *knows* the criterion values for those cases. The idea is to combine the predictors for maximal prediction for *future* samples" (p. 1306). Precision Efficacy Analysis for Regression has been shown through a line of research (Brooks, 1998a, 1998b; Brooks & Barcikowski, 1994, 1995, 1996) to be a viable method for this generalizability analysis.

The PEAR method appears to fill an important gap in the regression literature in that it recommends sample sizes for prediction based not only on the number of predictors in a study, but also on the size of the effect expected. Indeed, most sample size methods in other areas of statistics, including fixed model regression, consider effect size to be an essential part of the calculation. The PEAR method provides a means by which researchers can choose samples by setting a priori effect sizes, shrinkage tolerance, and precision efficacy levels. Brooks (1998a) and Brooks & Barcikowski (1995) have shown that prediction models produced using

appropriately large sample sizes will better estimate ρ_C^2. The most important argument for the PEAR method is that a model based on a proper sample size, as suggested by the PEAR method, will provide more reliable regression weights. Therefore, these models will predict better for future subjects because, ultimately, the efficiency of a prediction model depends not only on correlation statistics such as R^2 and R_C^2, but also on the stability of the regression coefficients used to calculate predicted scores.

From the relative efficiency statistics it would seem that the $PE = .80$ level used with the PEAR method usually would be most desirable. However, rather than rely on such a generalization, researchers must be consider the needs of each project. For example, at lower population ρ^2 effect sizes, the statistics based on the methods become rather close in absolute value. For example, at $\rho^2 = .10$ with three predictors, R_C^2 was .088 and SE_{b_j} averaged 0.05 for the $PE = .80$ level but $R_C^2 = .077$ with average $SE_{b_j} = 0.07$ for $PE = .60$. The $PE = .80$ level required 331 subjects to obtain its larger R_C^2, whereas the $PE = .60$ level only required 168 subjects to obtain a value that many researchers might find acceptable (Brooks, 1998b). Other researchers may determine, however, that the additional subjects recommended by the $PE = .80$ level are well worth the added precision efficacy. These dramatic differences in sample sizes must be balanced against the expected gain in precision and R_C^2, particularly at lower effect sizes. The sample size differences are not quite so striking at higher effect sizes, but still must be considered. For example, at $\rho^2 = .40$ and three predictors, the extra 28 subjects recommended by the $PE = .80$ ($N \geq 59$) level as compared to the $PE = .60$ level ($N \geq 31$) resulted in the more noticeable difference in average R_C^2 of .350 versus .294, respectively, and SE_{b_j} of 0.10 and 0.14, also respectively. Fortunately, thoughtful adjustments to the a priori precision efficacy or the shrinkage tolerance enable researchers to use the PEAR method to make such choices.

Some may argue that effect sizes required by the PEAR method are too difficult to determine—"if one knew the answer to that question one would not need to do the study. . ." (Schafer, 1993, p. 387)—but blind adherence to conventional subject-to-predictor ratios certainly cannot be better research practice. Further, research in the evolution of the PEAR method has determined that when expected R^2 overestimates the actual ρ^2 value by too much (e.g., based on an effect size too large or due to an inappropriate conventional rule), no regression sample size

method will recommend appropriate sample sizes for generalizability. For example, Brooks and Barcikowski (1995) found that when $R_E^2 = .25$ but $\rho^2 = .10$, precision efficacy rates were in the .47 to .50 range even for $PE = .80$. This reinforces the need for carefully chosen effect sizes in regression—as Schafer (1993) continued, ". . . but a value is needed anyway" (p. 387). When effect sizes are difficult to determine, pilot studies, meta-analyses, and careful interpretation of previous research play a critical role in the research process. Fortunately, because the PEAR method has performed well at a variety of effect sizes, numbers of predictors, shrinkage tolerance levels, and levels of multicollinearity, it seems to be well-suited to a variety of research situations.

These results are based on long-run expectations of the performance of the PEAR method. Berry (1993) has noted that "unbiasedness of OLS [ordinary least squares] estimators in no way ensures that an individual estimate of a regression parameter based on a single sample will equal its population value" (p. 18). Similarly, although the expected value of precision efficacy has been shown to be accurate in the long-run, any given sample size based on the PEAR method may not produce a PE value within the stringent accuracy range used in this study. However, results based on larger samples are less likely to vary, because larger samples generally result in smaller standard errors.

Therefore, developing a model with good precision efficacy should be considered only a first step in the model validation process. The use of mathematical cross-validity formulas does not supersede the need for the validation of regression models in other samples. The cross-validity formulas suggest how well a model should perform, but the safest way to determine that a model will generalize to future subjects is to test it with new data. Indeed, replication is basic to all science and is essential to confidence in both the reliability and the generalizability of results. Additionally, Darlington (1990) and Montgomery and Peck (1992) have expressed the importance not only of model validation but also of model adequacy, which requires residual analyses for violations of assumptions, searching for high leverage or overly influential observations, and other analyses that test the fit of the regression model to the available data. Darlington noted, however, that robustness to certain violations of assumptions continues to increase as sample size increases.

The use of mathematical cross-validity formulas does not supersede the need for the

validation of regression models in other samples. The cross-validity formulas suggest how well a model should perform, assuming that the sample from which it was derived was reasonably representative of the population; however, any given sample can deviate from what would be expected or representative. Further, no matter what the precision efficacy, a model that does not predict well in a derivation sample also probably will not predict well in any other samples. Finally, cross-validation does not depend upon the assumptions required for use of the cross-validity equations, thus providing a possible substitute when the assumptions are not met (Darlington, 1990; Wherry, 1975).

It is hoped that both the evidence presented and the simplicity of the PEAR method will encourage researchers to consider more carefully the issues of sample size, effect size, and generalizability for regression research. Because generalizability may be an even more important issue than statistical power in much regression research, an assessment technique such as Precision Efficacy Analysis for Regression appears beneficial to a more complete understanding of regression results.

References

Barcikowski, R. S. (1980). Regression Analysis. Unpublished manuscript, Ohio University at Athens.

Berry, W. D. (1993). Understanding regression assumptions (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-092). Thousand Oaks, CA: Sage.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. Technometrics, 37, 373-384.

Brooks, G. P. (1998a, October). Precision Efficacy Analysis for Regression. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL.

Brooks, G. P. (1998b). Precision efficacy analysis for regression: Development and justification of a new sample size method for multiple linear regression. Unpublished doctoral dissertation, Ohio University, Athens.

Brooks, G. P., & Barcikowski, R. S. (1994, April). A new sample size formula for regression. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 412247).

Brooks, G. P., & Barcikowski, R. S. (1995, October). Precision power method for selecting sample sizes. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 412246).

Brooks, G. P., & Barcikowski, R. S. (1996). Precision power and its application to the selection of regression sample sizes. Mid-Western Educational Researcher, 9(4), 10-17.

Browne, M. W. (1975). Predictive validity of a linear regression equation. British Journal of Mathematical and Statistical Psychology, 28, 79-87.

Cattin, P. (1980a). Estimation of the predictive power of a regression model. Journal of Applied Psychology, 65, 407-414.

Cattin, P. (1980b). Note on the estimation of the squared cross-validated multiple correlation of a regression model. Psychological Bulletin, 87, 63-65.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Darlington, R. B. (1990). Regression and linear models. New York: McGraw-Hill.

Darlington, R. B. (1996). Estimating the true accuracy of regression predictions. Mid-Western Educational Researcher, 9(4), 29-31.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology, 45, 265-282.

Drasgow, F., Dorans, N. J., & Tucker, L. R. (1979). Estimators of the squared cross-validity coefficient: A Monte Carlo investigation. Applied Psychological Measurement, 3, 387-399.

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample

size calculations. Psychological Bulletin, 106, 516-524.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26, 499-510.

Herzberg, P. A. (1969). The parameters of cross-validation. Psychometrika Monograph Supplement, 34 (2, Pt. 2).

Hinkle, D. E., & Oliver, J. D. (1983). How large should a sample be? A question with no simple answer? Or.... Educational and Psychological Measurement, 43, 1051-1060.

Huberty, C. J. (1989). Problems with stepwise methods—better alternatives. In B. Thompson (Ed.), Advances in social science methodology: A research annual (Vol. 1, pp. 43-70).Greenwich, CT: JAI.

Huberty, C. J. (1994). A note on interpreting an R^2 value. Journal of Educational and Behavioral Statistics, 19, 351-356.

Huberty, C. J., & Mourad, S. A. (1980). Estimation in multiple correlation/prediction. Educational and Psychological Measurement, 40, 101-112.

Jennrich, R. I. (1977). Stepwise regression. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), Mathematical methods for digital computers. Volume III: Statistical methods for digital computers (pp. 58-75). New York: Wiley-Interscience.

Kennedy, E. (1988). Estimation of the squared cross-validity coefficient in the context of best subset regression. Applied Psychological Measurement, 12, 231-237.

Kraemer, H. C., & Thiemann, S. (1987). How many subjects? Statistical power analysis in research. Newbury Park, CA: Sage.

Kromrey, J. D., & Hines, C. V. (1995). Use of empirical estimates of shrinkage in multiple regression: A caution. Educational and Psychological Measurement, 55, 901-925.

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. Communications of the ACM, 31, 742-749, 774.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). By design: Planning research on higher education. Cambridge, MA: Harvard University.

Lord, F. M. (1950). Efficiency of prediction when a regression equation from one sample is used in a new sample (Research Bulletin No. 50-40). Princeton, NJ: Educational Testing Service.

Mason, R. L., Gunst, R. F., & Webster, J. T. (1975). Regression analysis and problems of multicollinearity. Communications in Statistics, 4, 277-292.

Milton, S. (1986). A sample size formula for multiple regression studies. Public Opinion Quarterly, 50, 112-118.

Montgomery, D. C., & Peck, E. A. (1992). Introduction to linear regression analysis (2nd ed.). New York: John Wiley & Sons.

Mooney, C. Z. (1997). Monte Carlo simulation (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-116). Thousand Oaks, CA: Sage.

Morris, J. D. (1981). Updating the criterion for regression predictor variable selection. Educational and Psychological Measurement, 41, 777-780.

Nash, J. C. (1990). Compact numerical methods for computers: Linear algebra and function minimisation (2nd ed.). New York: Adam Hilger.

Nicholson, G. E. (1960). Prediction in future samples. In I. Olkin et al. (Eds.), Contributions to probability and statistics: Essays in honor of Harold Hotelling (pp. 322-330).

Palo Alto, CA: Stanford University.

Park, C. N., & Dudycha, A. L. (1974). A cross-validation approach to sample size determination for regression models. Journal of the American Statistical Association, 69, 214-218.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Lawrence Erlbaum Associates.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). Numerical recipes in Pascal: The art of scientific computing. New York: Cambridge University.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical recipes in FORTRAN: The art of scientific computing (2nd ed.). New York: Cambridge University.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlations: A clarification. Psychological Bulletin, 85, 1348-1351.

Rozeboom, W. W. (1981). The cross-validational accuracy of sample regressions. Journal of Educational Statistics, 6, 179-198.

Sawyer, R. (1982). Sample size and the accuracy of predictions made from multiple regression equations. Journal of Educational Statistics, 7, 91-104.

Schafer, W. D. (1993). Interpreting statistical significance and nonsignificance. Journal of Experimental Education, 61, 383-387.

Schmitt, N., Coyle, B. W., & Rauschenberger, J. (1977). A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psychological Bulletin, 84, 751-758.

SPSS Inc. (1996). SPSS® 7.0 statistical algorithms. Chicago: Authors.

Stein, C. (1960). Multiple regression. In I. Olkin et al. (Eds.), Contributions to probability and statistics: Essays in honor of Harold Hotelling (pp.425-443). Palo Alto, CA: Stanford University.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics (2nd ed.). New York: HarperCollins.

Uhl, N., & Eisenberg, T. (1970). Predicting shrinkage in the multiple correlation coefficient. Educational and Psychological Measurement, 30, 487-489.

Weisberg, S. (1985). Applied linear regression (2nd ed.). New York: John Wiley & Sons.

Wherry, R. J., Sr. (1975). Underprediction from overfitting: 45 years of shrinkage. Personnel Psychology, 28, 1-18.

Table 1

Sample Sizes at Two Levels of Expected Sample (R^2) and Four Predictors

Method	Assumed Population Squared Correlation $R_E^2 = .25$	$R_E^2 = .10$[a]
Cohen (1988) [$1 - \beta = .90$, $\alpha = .05$]	48	144
Darlington (1990) Precision Analysis[b]	166	230
Darlington (1990) Specific Conclusions	42	134
Gatsonis & Sampson (1989) [$1 - \beta = .90$, $\alpha = .05$]	55	165
Milton (1986) [$t = 2$, $\Delta r_j^2 = .02$, $\alpha = .05$]	155	185
Park & Dudycha (1974) [$\gamma = .90$][c]	93	173
PEAR method [$\epsilon = .22R_E^2$]	142	414
15:1 (Stevens, 1996)	60	60
30:1 (Pedhazur & Schmelkin, 1991)	120	120
$50 + 8p$ (Green, 1991)	82	82
Sawyer (1982) [$K = 1.05$]	55	55

[a] for Cohen (1988) and Gatsonis & Sampson (1989), actually $R_E = .30$.

[b] for $R_E^2 = .25$, lower confidence limit (LCL) is .16; for $R_E^2 = .10$, $LCL = .04$.

[c] for $R_E^2 = .25$, $\epsilon = .05$; for $R_E^2 = .10$, $\epsilon = .03$.

Table 2

Subjects per Variable[a] Sample Size Ratios from the PEAR Method and the *15:1* ratio

	Precision Efficacy (*PE*)			
ρ_E^2	.60	.70	.80	*15:1* ratio
.05	87.4	116.2	173.7	15.0
.10	41.9	55.5	82.8	15.0
.15	26.8	35.3	52.5	15.0
.20	19.2	25.2	37.4	15.0
.25	14.6	19.2	28.3	15.0
.30	11.6	15.1	22.2	15.0
.35	9.4	12.3	17.9	15.0
.40	7.8	10.1	14.6	15.0
.45	6.6	8.4	12.1	15.0
.50	5.5	7.1	10.1	15.0
.55	4.7	6.0	8.4	15.0
.60	4.0	5.0	7.1	15.0
.65	3.4	4.3	5.9	15.0
.70	2.9	3.6	4.9	15.0
.75	2.5	3.0	4.0	15.0

Note. Here, $\epsilon = \rho_E^2 - (PE - .1PS)\rho_E^2$, where $PS = 1 - PE$ and ρ_E^2 is the estimated population ρ^2 value. To calculate *N*, multiply the number of variables by the tabled value and round to the next larger integer if necessary

[a] number of variables is *(p+1)*, where *p* is the number of predictors.

Table 3

Stein-Darlington Cross-Validity Estimates based on Sample Sizes from Several Methods at Two Levels of Expected Sample Squared Multiple Correlation (R_E^2) and Four Predictors

Method	R_E^2 = .25			R_E^2 = .10		
	N	R_C^2	*PE*[a]	*N*	R_C^2	*PE*[a]
Cohen (1988)	48	.083	.33	144	.041	.41
Darlington (1990)[b]	166	.207	.83	230	.064	.64
Darlington (1990)[c]	42	.055	.22	134	.036	.36
Gatsonis & Sampson (1989)	55	.108	.43	165	.049	.49
Milton (1986)	155	.204	.82	185	.054	.54
Park & Dudycha (1974)	93	.171	.68	173	.051	.51
PEAR method [$\epsilon = .22\rho_E^2$]	142	.199	.80	414	.080	.80
PEAR method [$\epsilon = .33\rho_E^2$]	96	.174	.69	278	.070	.70
15:1 (Stevens, 1996)	60	.121	.49	60	-.054	.00
30:1 (Pedhazur & Schmelkin, 1991)	120	.190	.76	120	.028	.28
$50 + 8p$ (Green, 1991)	82	.159	.64	82	-.009	.00
Sawyer (1982) [$K = 1.05$]	55	.108	.43	55	-.070	.00

[a] *PE* here is calculated as R_C^2/ρ^2. [b] Precision Analysis. [c] Specific Conclusions.

Table 4
Average Standard Errors of the Standardized Coefficients (SE_{b_j}) for Three Predictors

Multicollinearity Condition	ρ^2	Method	*N*	SE_{b_1}	SE_{b_2}	SE_{b_3}
Orthogonal	.40	*PE = .80*	59	.102	.103	.094
		PE = .70	40	.126	.126	.118
		PE = .60	31	.147	.147	.136
		15:1 ratio	45	.119	.118	.109
	.25	*PE = .80*	113	.080	.080	.079
		PE = .70	77	.098	.099	.097
		PE = .60	59	.114	.113	.111
		15:1 ratio	45	.131	.132	.128
	.10	*PE = .80*	331	.052	.052	.050
		PE = .70	222	.064	.064	.062
		PE = .60	168	.074	.073	.071
		15:1 ratio	45	.146	.147	.143
Non	.40	*PE = .80*	59	.108	.108	.096
		PE = .70	40	.134	.135	.120
		PE = .60	31	.155	.155	.139
		15:1 ratio	45	.127	.126	.111
	.25	*PE = .80*	113	.139	.082	.136
		PE = .70	77	.170	.100	.166
		PE = .60	59	.195	.115	.193
		15:1 ratio	45	.228	.132	.223
	.10	*PE = .80*	331	.071	.066	.055
		PE = .70	222	.089	.083	.068
		PE = .60	168	.101	.095	.079
		15:1 ratio	45	.204	.189	.155
Moderate	.40	*PE = .80*	59	.202	.254*	.140
		PE = .70	40	.254	.312*	.173
		PE = .60	31	.295	.365*	.201
		15:1 ratio	45	.236	.293*	.160
	.25	*PE = .80*	113	.154	.213*	.146
		PE = .70	77	.189	.260*	.177
		PE = .60	59	.218	.302*	.210
		15:1 ratio	45	.252	.349*	.239
	.10	*PE = .80*	331	.114	.151*	.090
		PE = .70	222	.140	.187*	.113
		PE = .60	168	.160	.213*	.128
		15:1 ratio	45	.327	.436*	.260
Extensive	.40	*PE = .80*	59	.183	.264*	.308*
		PE = .70	40	.228	.327*	.382*
		PE = .60	31	.264	.387*	.453*
		15:1 ratio	45	.212	.308*	.357*
	.25	*PE = .80*	113	.129	.381*	.407*
		PE = .70	77	.158	.466*	.499*
		PE = .60	59	.179	.537*	.573*
		15:1 ratio	45	.209	.631*	.672*
	.10	*PE = .80*	331	.128*	.124*	.065
		PE = .70	222	.156*	.152*	.080
		PE = .60	168	.180*	.176*	.093
		15:1 ratio	45	.363*	.352*	.185

Note. SE_{b_j} approximates *RMSE* when estimate is unbiased as is β_j.
* indicates predictor with *VIF*> 5.0 (i.e., involved in multicollinearity).

Table 5
Average Standard Errors of the Standardized Coefficients (SE_{b_j}) for Seven Predictors

Multicollinearity Condition	ρ^2	Method	*N*	SE_{b_1}	SE_{b_2}	SE_{b_3}	SE_{b_4}	SE_{b_5}	SE_{b_6}	SE_{b_7}
Orthogonal	40	*PE* = .80	117	.074	.073	.073	.068	.073	.074	.074
		PE = .70	81	.091	.089	.089	.083	.089	.091	.091
		PE = .60	63	.105	.103	.102	.098	.105	.104	.105
		15:1 ratio	105	.079	.079	.077	.071	.078	.078	.078
	25	*PE* = .80	226	.058	.058	.056	.056	.058	.058	.058
		PE = .70	153	.071	.072	.069	.069	.071	.072	.071
		PE = .60	117	.084	.083	.080	.080	.082	.083	.082
		15:1 ratio	105	.087	.089	.085	.083	.088	.086	.087
	10	*PE* = .80	663	.036	.037	.037	.036	.037	.037	.037
		PE = .70	444	.044	.044	.045	.045	.046	.046	.045
		PE = .60	335	.051	.052	.053	.052	.052	.053	.052
		15:1 ratio	105	.093	.095	.096	.095	.096	.097	.096
Non	40	*PE* = .80	117	.100	.102	.100	.097	.109	.091	.081
		PE = .70	81	.123	.124	.123	.119	.135	.111	.099
		PE = .60	63	.142	.144	.141	.138	.156	.126	.116
		15:1 ratio	105	.105	.108	.106	.103	.117	.096	.087
	25	*PE* = .80	226	.070	.085	.070	.064	.090	.071	.079
		PE = .70	153	.087	.105	.086	.078	.109	.086	.098
		PE = .60	117	.099	.121	.098	.089	.127	.099	.113
		15:1 ratio	105	.106	.129	.105	.094	.135	.106	.120
	10	*PE* = .80	663	.043	.042	.050	.061	.052	.057	.054
		PE = .70	444	.053	.052	.060	.075	.065	.069	.066
		PE = .60	335	.060	.061	.071	.086	.075	.080	.076
		15:1 ratio	105	.111	.111	.128	.158	.137	.148	.138
Moderate	40	*PE* = .80	117	.192*	.137	.141	.177*	.154	.094	.130
		PE = .70	81	.236*	.170	.174	.219*	.188	.116	.158
		PE = .60	63	.270*	.191	.200	.249*	.215	.132	.182
		15:1 ratio	105	.203*	.146	.151	.188*	.161	.099	.136
	25	*PE* = .80	226	.129	.089	.130*	.079	.080	.074	.180*
		PE = .70	153	.159	.109	.160*	.097	.099	.092	.220*
		PE = .60	117	.184	.126	.187*	.113	.114	.107	.258*
		15:1 ratio	105	.196	.134	.197*	.119	.121	.110	.273*
	10	*PE* = .80	663	.086*	.043	.098*	.083	.060	.047	.041
		PE = .70	444	.103*	.052	.120*	.101	.072	.058	.051
		PE = .60	335	.121*	.061	.139*	.118	.084	.066	.059
		15:1 ratio	105	.222*	.110	.256*	.216	.154	.123	.107
Extensive	40	*PE* = .80	117	.118	.131	.166*	.168*	.256*	.228*	.132
		PE = .70	81	.143	.161	.199*	.204*	.306*	.273*	.158
		PE = .60	63	.167	.187	.233*	.236*	.359*	.318*	.184
		15:1 ratio	105	.125	.141	.175*	.178*	.269*	.239*	.138
	25	*PE* = .80	452	.093	.168*	.147*	.150*	.097	.121	.147*
		PE = .70	307	.113	.207*	.181*	.184*	.118	.150	.179*
		PE = .60	234	.131	.237*	.207*	.213*	.139	.173	.205*
		15:1 ratio	105	.138	.254*	.222*	.227*	.147	.186	.221*
	10	*PE* = .80	663	.153*	.136*	.083	.106*	.063	.047	.207*
		PE = .70	444	.185*	.164*	.101	.129*	.076	.058	.250*
		PE = .60	335	.213*	.187*	.114	.150*	.087	.066	.287*
		15:1 ratio	105	.390*	.348*	.213	.272*	.159	.123	.526*

Note. SE_{b_j} approximates *RMSE* when estimate is unbiased as is β_j.
* indicates predictor with *VIF* > 5.0 (i.e., involved in multicollinearity).

Table 6
Average Relative Efficiency of the Standardized Coefficients Across Predictors

ρ^2	p	Method Comparison	Orthogonal	Non	Moderate	Extensive
40	3	*RMSE*(.80) / *RMSE*(.70)	80.8	80.2	80.6	80.5
		RMSE(.80) / *RMSE*(.60)	69.5	69.5	69.2	68.5
		RMSE(.70) / *RMSE*(.60)	86.1	86.6	85.9	85.1
	7	*RMSE*(.80) / *RMSE*(.70)	81.7	81.6	81.3	82.9
		RMSE(.80) / *RMSE*(.60)	70.5	70.6	71.2	71.1
		RMSE(.70) / *RMSE*(.60)	86.3	86.6	87.6	85.8
	11	*RMSE*(.80) / *RMSE*(.70)	81.4	81.8	81.6	80.5
		RMSE(.80) / *RMSE*(.60)	70.7	70.4	70.8	70.5
		RMSE(.70) / *RMSE*(.60)	86.8	86.1	86.7	87.6
	15	*RMSE*(.80) / *RMSE*(.70)	81.7	81.5	80.4	81.9
		RMSE(.80) / *RMSE*(.60)	70.7	70.6	69.8	70.7
		RMSE(.70) / *RMSE*(.60)	86.5	86.7	86.9	86.3
25	3	*RMSE*(.80) / *RMSE*(.70)	81.3	81.9	82.0	81.7
		RMSE(.80) / *RMSE*(.60)	70.7	71.0	70.2	71.3
		RMSE(.70) / *RMSE*(.60)	87.0	86.7	85.7	87.4
	7	*RMSE*(.80) / *RMSE*(.70)	81.2	81.5	81.2	81.6
		RMSE(.80) / *RMSE*(.60)	70.0	71.0	69.9	70.7
		RMSE(.70) / *RMSE*(.60)	86.2	87.1	86.1	86.6
	11	*RMSE*(.80) / *RMSE*(.70)	81.4	81.6	81.6	81.4
		RMSE(.80) / *RMSE*(.60)	70.5	70.8	70.6	71.1
		RMSE(.70) / *RMSE*(.60)	86.6	86.8	86.5	87.3
	15	*RMSE*(.80) / *RMSE*(.70)	81.8	81.2	81.0	81.4
		RMSE(.80) / *RMSE*(.60)	71.2	70.6	70.2	70.5
		RMSE(.70) / *RMSE*(.60)	87.0	86.9	86.8	86.5
10	3	*RMSE*(.80) / *RMSE*(.70)	81.0	80.1	80.6	81.6
		RMSE(.80) / *RMSE*(.60)	70.6	69.8	70.8	70.5
		RMSE(.70) / *RMSE*(.60)	87.2	87.2	87.9	86.4
	7	*RMSE*(.80) / *RMSE*(.70)	81.9	81.6	82.1	82.4
		RMSE(.80) / *RMSE*(.60)	70.4	70.5	70.6	72.0
		RMSE(.70) / *RMSE*(.60)	86.0	86.4	86.0	87.4
	11	*RMSE*(.80) / *RMSE*(.70)	81.1	81.9	81.8	81.7
		RMSE(.80) / *RMSE*(.60)	70.4	70.9	71.2	70.6
		RMSE(.70) / *RMSE*(.60)	86.8	86.6	87.1	86.5
	15	*RMSE*(.80) / *RMSE*(.70)	81.0	80.9	81.7	81.1
		RMSE(.80) / *RMSE*(.60)	70.2	70.4	70.7	70.7
		RMSE(.70) / *RMSE*(.60)	86.6	87.1	86.5	87.2

Note. SE_{b_j} approximates *RMSE* when estimate is unbiased as is β_j.

Table 7
Average Precision Efficacy (*PE*) for Orthogonal Stepwise Analyses as Compared to Orthogonal Full Model

ρ^2	Method	*p*	N	k	PE	$PE_{R^2_{C(p)}}$	$PE_{R^2_{C(k)}}$
.40	PE = .80	3	59	2.492	.802	.783	.828
		7	117	3.528	.803	.783	.892
		11	176	4.074	.806	.785	.921
		15	234	4.871	.805	.784	.931
	PE = .70	3	40	2.088	.690	.634	.753
		7	81	3.166	.714	.666	.856
		11	121	3.840	.718	.674	.891
		15	161	4.372	.719	.670	.909
	PE = .60	3	31	1.780	.597	.508	.688
		7	63	2.894	.629	.546	.824
		11	94	3.677	.636	.561	.864
		15	125	4.071	.640	.556	.891
	15:1 *ratio*	3	45	2.213	.729	.685	.780
		7	105	3.422	.780	.755	.883
		11	165	4.003	.791	.767	.916
		15	225	4.821	.798	.774	.929
.25	PE = .80	3	59	2.712	.800	.788	.815
		7	117	3.595	.802	.786	.889
		11	176	6.034	.805	.789	.885
		15	234	4.695	.803	.785	.931
	PE = .70	3	40	2.319	.698	.653	.742
		7	81	3.220	.708	.669	.848
		11	121	5.365	.715	.677	.846
		15	161	4.303	.718	.679	.908
	PE = .60	3	31	1.980	.605	.531	.676
		7	63	2.947	.621	.553	.813
		11	94	4.858	.634	.563	.815
		15	125	4.042	.637	.569	.886
	15:1 *ratio*	3	45	1.626	.496	.400	.597
		7	105	2.839	.581	.498	.797
		11	165	4.724	.612	.530	.808
		15	225	3.995	.622	.549	.882
.10	PE = .80	3	59	2.126	.801	.787	.848
		7	117	4.133	.803	.790	.875
		11	176	6.132	.803	.793	.882
		15	234	6.544	.803	.789	.906
	PE = .70	3	40	1.829	.697	.659	.787
		7	81	3.500	.711	.673	.836
		11	121	5.591	.714	.687	.840
		15	161	5.689	.715	.680	.877
	PE = .60	3	31	1.631	.600	.540	.724
		7	63	3.062	.622	.554	.804
		11	94	5.011	.628	.571	.806
		15	125	5.058	.634	.570	.855
	15:1 *ratio*	3	45	0.728	.169	.111	.331
		7	105	1.597	.183	.088	.558
		11	165	2.572	.187	.073	.631
		15	225	2.907	.188	.062	.731

Note. PE is the precision efficacy for the full model; $PE_{R^2_{C(p)}}$ represents precision efficacy for the stepwise model calculated using R^2_C with the number of predictors in the full model; $PE_{R^2_{C(k)}}$ represents precision efficacy for the stepwise model calculated using R^2_C based on the number of predictors in the final stepwise model.

Table 8

Average Precision Efficacy for Stepwise Solution with Seven Predictors in Full Model

ρ^2	Method	Orthogonal	Non	Moderate	Extensive
.40	PE = .80	.783	.775	.699	.627
	PE = .70	.666	.629	.502	.429
	PE = .60	.546	.484	.355	.292
	15:1 *ratio*	.755	.738	.651	.571
.25	PE = .80	.786	.783	.740	.714
	PE = .70	.669	.661	.562	.509
	PE = .60	.553	.538	.410	.356
	15:1 *ratio*	.498	.481	.346	.296
.10	PE = .80	.790	.786	.783	.765
	PE = .70	.673	.665	.664	.636
	PE = .60	.554	.543	.542	.508
	15:1 *ratio*	.088	.080	.082	.074

Note. Precision Efficacy for the stepwise solution is $PE_{R^2_{(\Psi)}}$ based on the Stein-Darlington formula using the total number of predictors in the full model (p).

Table 9
Average Cross-Validity Estimates and their *RMSE* for the Several Multicollinearity Conditions

ρ^2	*p*	Method	Orthogonal	Non	Moderate	Extensive
.40	3	*PE* = .80	.337 (.102)	.337 (.109)	.242 (.117)	.300 (.109)
		PE = .70	.288 (.126)	.296 (.129)	.183 (.142)	.232 (.134)
		PE = .60	.247 (.138)	.259 (.141)	.146 (.153)	.186 (.147)
		15:1 *ratio*	.302 (.117)	.310 (.123)	.200 (.135)	.254 (.127)
	7	*PE* = .80	.327 (.088)	.323 (.085)	.269 (.151)	.208 (.122)
		PE = .70	.289 (.114)	.270 (.111)	.199 (.177)	.151 (.150)
		PE = .60	.249 (.139)	.218 (.135)	.151 (.194)	.113 (.176)
		15:1 *ratio*	.319 (.094)	.308 (.093)	.251 (.157)	.191 (.161)
	11	*PE* = .80	.327 (.081)	.334 (.071)	.306 (.092)	.290 (.080)
		PE = .70	.290 (.115)	.287 (.102)	.237 (.096)	.234 (.102)
		PE = .60	.251 (.147)	.240 (.129)	.181 (.109)	.183 (.123)
		15:1 *ratio*	.320 (.088)	.328 (.075)	.297 (.088)	.282 (.083)
	15	*PE* = .80	.326 (.073)	.315 (.107)	.323 (.074)	.322 (.189)
		PE = .70	.285 (.107)	.268 (.131)	.280 (.105)	.278 (.178)
		PE = .60	.244 (.141)	.222 (.158)	.241 (.134)	.233 (.188)
		15:1 *ratio*	.322 (.076)	.311 (.108)	.319 (.079)	.318 (.187)
.25	3	*PE* = .80	.214 (.069)	.218 (.124)	.176 (.068)	.203 (.107)
		PE = .70	.189 (.078)	.196 (.122)	.132 (.077)	.168 (.107)
		PE = .60	.196 (.087)	.177 (.123)	.102 (.083)	.140 (.106)
		15:1 *ratio*	.139 (.095)	.153 (.126)	.078 (.086)	.112 (.108)
	7	*PE* = .80	.206 (.056)	.203 (.062)	.179 (.113)	.172 (.114)
		PE = .70	.183 (.074)	.178 (.076)	.142 (.132)	.126 (.101)
		PE = .60	.159 (.089)	.154 (.084)	.111 (.124)	.095 (.106)
		15:1 *ratio*	.148 (.095)	.142 (.088)	.098 (.129)	.082 (.103)
	11	*PE* = .80	.208 (.045)	.207 (.058)	.179 (.103)	.140 (.047)
		PE = .70	.183 (.059)	.180 (.065)	.125 (.094)	.095 (.056)
		PE = .60	.159 (.071)	.147 (.073)	.086 (.086)	.066 (.065)
		15:1 *ratio*	.152 (.073)	.140 (.074)	.077 (.085)	.060 (.067)
	15	*PE* = .80	.203 (.049)	.194 (.043)	.185 (.066)	.201 (.044)
		PE = .70	.181 (.068)	.167 (.061)	.157 (.069)	.176 (.063)
		PE = .60	.157 (.088)	.140 (.082)	.128 (.080)	.150 (.083)
		15:1 *ratio*	.153 (.092)	.135 (.086)	.123 (.082)	.146 (.086)
.10	3	*PE* = .80	.084 (.031)	.077 (.031)	.070 (.033)	.074 (.032)
		PE = .70	.075 (.038)	.067 (.040)	.059 (.041)	.063 (.040)
		PE = .60	.067 (.044)	.058 (.046)	.051 (.046)	.054 (.045)
		15:1 *ratio*	.030 (.057)	.025 (.056)	.020 (.051)	.024 (.054)
	7	*PE* = .80	.084 (.023)	.082 (.024)	.081 (.023)	.074 (.024)
		PE = .70	.074 (.028)	.072 (.029)	.071 (.028)	.064 (.031)
		PE = .60	.064 (.033)	.062 (.034)	.062 (.034)	.055 (.039)
		15:1 *ratio*	.019 (.047)	.017 (.047)	.018 (.047)	.016 (.051)
	11	*PE* = .80	.084 (.019)	.087 (.032)	.071 (.021)	.073 (.022)
		PE = .70	.076 (.024)	.076 (.031)	.059 (.022)	.057 (.025)
		PE = .60	.065 (.028)	.061 (.031)	.049 (.025)	.042 (.027)
		15:1 *ratio*	.014 (.036)	.009 (.029)	.008 (.029)	.005 (.027)
	15	*PE* = .80	.083 (.017)	.080 (.018)	.076 (.021)	.068 (.018)
		PE = .70	.073 (.023)	.069 (.021)	.065 (.021)	.058 (.020)
		PE = .60	.063 (.028)	.059 (.025)	.055 (.026)	.049 (.026)
		15:1 *ratio*	.011 (.054)	.010 (.050)	.009 (.050)	.007 (.050)

Note. Standard deviations in parentheses. Cross-validity for the stepwise solution is represented by $R^2_{C(p)}$, as calculated by the Stein-Darlington formula using the number of predictors in the full model (*p*).

Table 10
Average Standard Errors of the Coefficients (SE_b) for Seven Predictor Models from the Stepwise Analyses for All Multicollinearity Conditions

Multicollinearity Condition	ρ^2	Method	*N*	SE_{b_1}	SE_{b_2}	SE_{b_3}	SE_{b_4}	SE_{b_5}	SE_{b_6}	SE_{b_7}
Orthogonal	40	*PE* = .80	117	.072	.073	.073	.073	.072	.072	.072
		PE = .70	81	.086	.087	.088	.088	.086	.086	.086
		PE = .60	63	.097	.099	.099	.101	.098	.097	.098
		15:1 *ratio*	105	.076	.077	.077	.077	.076	.076	.076
	25	*PE* = .80	226	.058	.058	.058	.058	.058	.058	.058
		PE = .70	153	.070	.070	.071	.071	.070	.070	.070
		PE = .60	117	.080	.080	.081	.081	.080	.080	.080
		15:1 *ratio*	105	.084	.085	.086	.086	.085	.084	.085
	10	*PE* = .80	663	.037	.037	.037	.037	.037	.037	.037
		PE = .70	444	.045	.045	.045	.045	.045	.045	.045
		PE = .60	335	.052	.052	.052	.052	.052	.052	.052
		15:1 *ratio*	105	.093	.092	.092	.092	.092	.093	.093
Non	40	*PE* = .80	117	.088	.083	.092	.085	.100	.079	.080
		PE = .70	81	.105	.098	.107	.100	.116	.093	.095
		PE = .60	63	.118	.109	.119	.112	.128	.103	.107
		15:1 *ratio*	105	.093	.087	.097	.090	.104	.083	.084
	25	*PE* = .80	226	.067	.064	.065	.063	.070	.061	.068
		PE = .70	153	.081	.077	.077	.076	.082	.073	.082
		PE = .60	117	.092	.087	.087	.086	.093	.083	.094
		15:1 *ratio*	105	.097	.091	.091	.091	.097	.088	.099
	10	*PE* = .80	663	.039	.040	.046	.045	.045	.042	.040
		PE = .70	444	.047	.049	.056	.052	.053	.051	.048
		PE = .60	335	.054	.056	.064	.059	.061	.058	.055
		15:1 *ratio*	105	.094	.096	.104	.099	.102	.100	.095
Moderate	40	*PE* = .80	117	.145*	.106	.126	.118*	.128	.082	.112
		PE = .70	81	.157*	.118	.148	.124*	.140	.099	.128
		PE = .60	63	.164*	.128	.165	.132*	.149	.112	.137
		15:1 *ratio*	105	.148*	.110	.132	.119*	.132	.086	.117
	25	*PE* = .80	226	.108	.084	.083*	.071	.075	.069	.156*
		PE = .70	153	.117	.090	.094*	.083	.090	.082	.167*
		PE = .60	117	.124	.110	.102*	.092	.100	.091	.171*
		15:1 *ratio*	105	.127	.115	.106*	.097	.105	.096	.174*
	10	*PE* = .80	663	.041*	.039	.059*	.041	.042	.040	.040
		PE = .70	444	.050*	.048	.068*	.049	.051	.048	.049
		PE = .60	335	.057*	.055	.076*	.056	.058	.056	.056
		15:1 *ratio*	105	.098*	.096	.122*	.098	.101	.096	.097
Extensive	40	*PE* = .80	117	.095	.114	.136*	.097*	.203*	.129*	.094
		PE = .70	81	.108	.125	.138*	.106*	.195*	.128*	.105
		PE = .60	63	.120	.132	.136*	.117*	.188*	.134*	.116
		15:1 *ratio*	105	.098	.117	.138*	.099*	.202*	.127*	.096
	25	*PE* = .80	452	.081	.093*	.089*	.087*	.071	.090	.097*
		PE = .70	307	.093	.107*	.094*	.097*	.083	.104	.106*
		PE = .60	234	.100	.118*	.100*	.105*	.094	.115	.116*
		15:1 *ratio*	225	.102	.123*	.103*	.107*	.098	.119	.121*
	10	*PE* = .80	663	.057*	.062*	.056	.044*	.047	.044	.075*
		PE = .70	444	.064*	.071*	.066	.050*	.056	.053	.082*
		PE = .60	335	.071*	.080*	.073	.057*	.064	.061	.091*
		15:1 *ratio*	105	.110*	.134*	.112	.100*	.104	.105	.143*

* indicates predictor with VIF > 5.0 (i.e., involved in multicollinearity)

Table 11
Relative Efficiency for Seven Predictor Models from the Stepwise Analyses

Multicollinearity Condition	ρ^2	Method Comparison	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Orthogonal	.40	*RMSE*(.80) / *RMSE*(.70)	84	84	83	83	84	84	84
		RMSE(.70) / *RMSE*(.60)	89	88	89	87	88	89	88
	.25	*RMSE*(.80) / *RMSE*(.70)	83	83	82	82	83	83	83
		RMSE(.70) / *RMSE*(.60)	88	88	88	88	88	88	88
	.10	*RMSE*(.80) / *RMSE*(.70)	82	82	82	82	82	82	82
		RMSE(.70) / *RMSE*(.60)	87	87	87	87	87	87	87
Non	.40	*RMSE*(.80) / *RMSE*(.70)	84	85	86	85	86	85	84
		RMSE(.70) / *RMSE*(.60)	89	90	90	89	91	90	89
	.25	*RMSE*(.80) / *RMSE*(.70)	83	83	84	83	85	84	83
		RMSE(.70) / *RMSE*(.60)	88	89	89	88	88	88	87
	.10	*RMSE*(.80) / *RMSE*(.70)	83	82	82	87	85	82	83
		RMSE(.70) / *RMSE*(.60)	87	88	88	88	87	88	87
Moderate	.40	*RMSE*(.80) / *RMSE*(.70)	92[a]	90	85	95[a]	91	83	88
		RMSE(.70) / *RMSE*(.60)	96[a]	92	90	94[a]	94	88	93
	.25	*RMSE*(.80) / *RMSE*(.70)	92	85	88[a]	86	83	84	93[a]
		RMSE(.70) / *RMSE*(.60)	94	90	92[a]	90	90	90	98[a]
	.10	*RMSE*(.80) / *RMSE*(.70)	82[a]	81	87[a]	84	82	83	82
		RMSE(.70) / *RMSE*(.60)	88[a]	87	89[a]	88	88	86	88
Extensive	.40	*RMSE*(.80) / *RMSE*(.70)	88	91	99[a]	92[a]	104[a]	100[a]	90
		RMSE(.70) / *RMSE*(.60)	90	95	100[a]	91[a]	104[a]	96[a]	91
	.25	*RMSE*(.80) / *RMSE*(.70)	87	87[a]	95[a]	90[a]	86	87	92[a]
		RMSE(.70) / *RMSE*(.60)	93	91[a]	94[a]	92[a]	88	90	91[a]
	.10	*RMSE*(.80) / *RMSE*(.70)	89[a]	87[a]	85	88[a]	84	83	91[a]
		RMSE(.70) / *RMSE*(.60)	90[a]	89[a]	90	88[a]	88	87	90[a]

[a] indicates predictor with *VIF* > 5.0 (i.e., involved in multicollinearity)

Appendix A

Derivation of the PEAR Method for Sample Size Selection

Start with the Lord formula, as presented by Uhl & Eisenberg (1970):

$$R_C^2 = 1 - \frac{N + p + 1}{N - p - 1}(1 - R^2)$$

Multiplying both sides by (N-p-1) yields:

$$(N - p - 1)(R_C^2) = (N - p - 1) - (N + p + 1)(1 - R^2)$$

Expanding the quantities gives:

$$NR_C^2 - pR_C^2 - R_C^2 = N - p - 1 - N - p - 1 + NR^2 + pR^2 + R^2$$

and grouping and subtracting gives:

$$NR_C^2 - NR^2 = pR_C^2 + R_C^2 - p - 1 - p - 1 + pR^2 + R^2$$

By factoring the terms:

$$N(R_C^2 - R^2) = p(R_C^2 - 2 + R^2) + 1(R_C^2 - 2 + R^2)$$

And therefore

$$N(R_C^2 - R^2) = (p + 1)(R_C^2 - 2 + R^2)$$

Multiplying both sides by (-1) and then dividing both sides by $(R^2 - R_C^2)$ gives:

$$N = (p + 1)\frac{(2 - R^2 - R_C^2)}{(R^2 - R_C^2)}$$

Let $\epsilon = R^2 - R_C^2$ and therefore $R_C^2 = R^2 - \epsilon$:

$$N = (p + 1)\frac{(2 - R^2 - (R^2 - \epsilon))}{\epsilon}$$

Finally,

$$N = (p + 1)\frac{(2 - 2R^2 + \epsilon)}{\epsilon}$$

Appendix B

Stem-and-Leaf Plots of the Precision Efficacy Accuracy of Several Sample Size Methods

These plots were adapted from Brooks and Barcikowski (1995). The accuracy criterion used for these results was $.75 \leq PE \leq .85$. Leaves which represent accurate results have been boldfaced and underlined. For every plot, the stem width is 0.10. Each leaf represents one case.

PEAR Method (Brooks & Barcikowski, 1995)

Frequency	Stem &	Leaf
.00	0 .	
.00	1 .	
.00	2 .	
.00	3 .	
.00	4 .	
.00	5 .	
.00	6 .	
1.00	7 .	**9**
19.00	8 .	**00111111111112222222**
.00	9 .	

50 + 8p conventional rule (Green, 1991)

Frequency	Stem &	Leaf
.00	0 .	
2.00	1 .	06
2.00	2 .	39
1.00	3 .	4
.00	4 .	
2.00	5 .	18
2.00	6 .	58
1.00	7 .	2
4.00	8 .	**14**89
6.00	9 .	134667

Park and Dudycha (1974)

Frequency	Stem &	Leaf
.00	0 .	
.00	1 .	
.00	2 .	
.00	3 .	
1.00	4 .	4
5.00	5 .	11349
2.00	6 .	68
3.00	7 .	01**9**
6.00	8 .	**455**679
3.00	9 .	000

15:1 subject-to-predictor ratio (Stevens, 1996)

Frequency	Stem &	Leaf
.00	0 .	
5.00	1 .	56777
.00	2 .	
.00	3 .	
2.00	4 .	08
2.00	5 .	39
1.00	6 .	1
1.00	7 .	**9**
4.00	8 .	**1355**
5.00	9 .	01445

Sawyer (1982)

Frequency	Stem &	Leaf
.00	0 .	
.00	1 .	
.00	2 .	
1.00	3 .	9
6.00	4 .	022368
3.00	5 .	011
5.00	6 .	23455
1.00	7 .	**9**
4.00	8 .	**0112**
.00	9 .	

Cohen (1988)

Frequency	Stem &	Leaf
1.00	0 .	7
3.00	1 .	169
2.00	2 .	01
5.00	3 .	02468
6.00	4 .	034588
3.00	5 .	025
.00	6 .	
.00	7 .	
.00	8 .	
.00	9 .	

30:1 N-to-p (Pedhazur & Schmelkin, 1991)

Frequency	Stem &	Leaf
.00	0 .	
.00	1 .	
.00	2 .	
3.00	3 .	169
2.00	4 .	37
.00	5 .	
1.00	6 .	9
3.00	7 .	469
1.00	8 .	**0**
10.00	9 .	0122277777

Gatsonis and Sampson (1989)

Frequency	Stem &	Leaf
.00	0 .	
.00	1 .	
3.00	2 .	135
3.00	3 .	667
3.00	4 .	299
5.00	5 .	34668
4.00	6 .	1156
2.00	7 .	04
.00	8 .	
.00	9 .	

BEST COPY AVAILABLE

Appendix C

Correlation Matrices for ρ^2 = .25 at Three and Seven Predictors

Number of Predictors	Multicollinearity Condition	ρ^2	y	x_1	x_2	x_3	x_4	x_5	x_6	
3	Orthogonal	.25	x_1	.257						
			x_2	.257	.000					
			x_3	.343	.000	.000				
	Non	.25	x_1	.257						
			x_2	.257	-.206					
		*	x_3	.343	.800	-.277				
	Moderate	.25	x_1	.257						
			x_2*	.257	.709					
			x_3	.343	.131	.683				
	Extensive	.25	x_1	.257						
			x_2*	.257	.680					
			x_3*	.343	.741	.976				
7	Orthogonal	.25	x_1	044						
			x_2	038	.000					
			x_3	318	000	.000				
			x_4	325	.000	.000	.000			
			x_5	148	000	.000	.000	.000		
			x_6	016	.000	.000	.000	.000	.000	
			x_7	132	.000	.000	.000	.000	000	.000
	Non	.25	x_1	044						
			x_2	038	.343					
			x_3	.318	.312	.188				
			x_4	325	.372	.019	.050			
			x_5	148	.340	.637	422	.180		
			x_6	016	027	.260	125	.017	.117	
			x_7	132	.136	.105	.464	.214	.364	.438
	Moderate	.25	x_1	044						
			x_2	038	.198					
			x_3*	318	.021	.510				
			x_4	325	024	.392	475			
			x_5	148	235	.557	.126	.304		
			x_6	016	.493	.427	265	.245	.263	
			x_7*	132	.682	.578	.614	.201	.410	.505
	Extensive	.25	x_1	044						
			x_2*	038	-.445					
			x_3*	318	.082	.095				
			x_4*	325	.063	.425	491			
			x_5	148	.533	-.195	.463	.043		
			x_6	016	.449	-.346	.152	.408	.387	
			x_7*	132	.117	.420	.658	.318	.548	.165

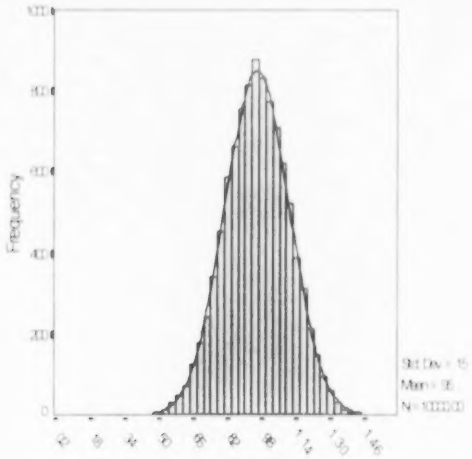
* indicates predictor with *VIF* > 5.0 (i.e., involved in multicollinearity)

Appendix D

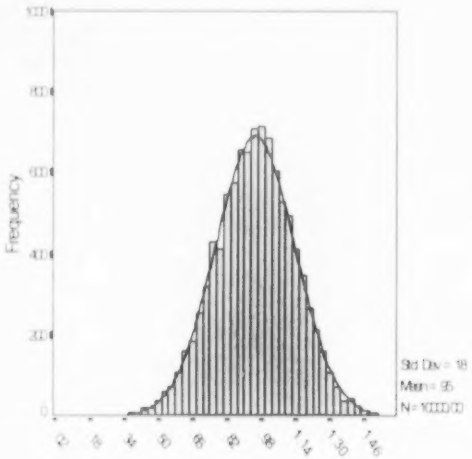
Histograms of the Distributions of the Coefficient for One Predictor at Effect Size ρ^2 = .25

These figures were created from data collected for each of the 10,000 samples at effect size ρ^2 = .25 with seven predictors in the extensive multicollinearity condition. Each figure represents a different precision efficacy level or the *15:1* ratio. In each of the following figures, a curve that represents the normal distribution is superimposed on the distribution of the fourth regression coefficient from the given set of conditions (compare to Table 5).

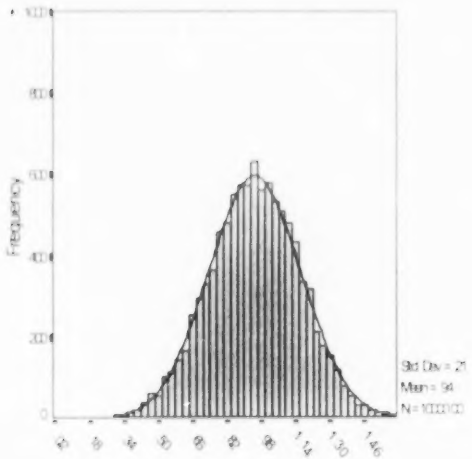
a priori *PE* = .80



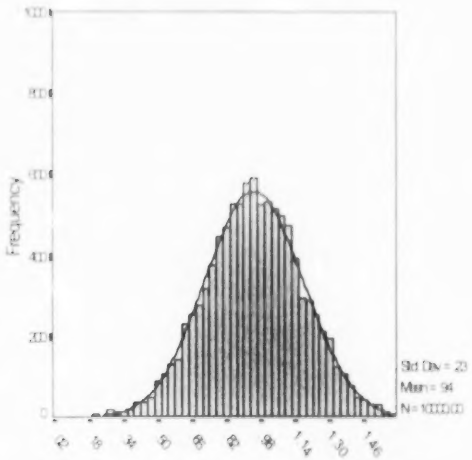
a priori *PE* = .70



a priori *PE* = .60



15:1 subject-to-predictor ratio



BEST COPY AVAILABLE